Survivorship Bias and Mutual Fund Performance: Relevance, Significance, and Methodical Differences*

MARTIN ROHLEDER[1], HENDRIK SCHOLZ[2] and MARCO WILKENS[3]
[1]*Catholic University of Eichstaett-Ingolstadt;* [2]*University of Erlangen-Nuernberg;*
[3]*University of Augsburg*

Abstract. This is the first paper systematically calculating, testing and explaining different definitions of the survivorship bias in fund performance. We document that the survival-performance-relation is stronger for small funds and we find significant under-performance of non-survivors but no significant out-performance of new funds. Survivorship bias is still a problem as well in other fields of research, e.g., in countries where survivorship bias-free data is not available and because certain methods require truncated data. This paper provides guidance on how to deal with and reduce survivorship bias in empirical studies.

JEL Classification: G11, G12

1. Introduction and Literature

Survivorship bias and its explanations are often addressed in mutual fund literature in context with data selection. It arises during measurement of the performance of a fund portfolio that includes only surviving funds. Generally, this leads to overestimating the performance of funds because the predominant reason for fund disappearance is inferior performance (e.g., Malkiel, 1995; Brown and Goetzmann, 1995; Elton et al., 1996). For example, our analysis shows a survivorship biased Jensen alpha of +48 basis points per year for the equal-weighted US domestic equity mutual fund market from 1993 through 2006, suggesting that the average fund out-performed the passive benchmark on a risk adjusted basis. The corresponding

* We are grateful for helpful comments and suggestions by Manuel Ammann, Wayne Ferson, William Goetzmann (Editor), Marc Guertler, Cesario Mateus, Steffen Meyer, Manuel Rocha Armada, Ray Sturm, Scott Yonker, an anonymous referee, and participants of the 2008 MFA meeting in San Antonio, the 2008 SWFA Meeting in Houston, the 2008 SGF Conference in Zurich, the 2008 EFA Meeting in St. Pete Beach, the 2008 FMA European Conference in Prague, the 2008 DGF Meeting in Muenster, and the 2008 SFA Meeting in Key West. We are responsible for any remaining errors.

unbiased Jensen alpha is −109 basis points per year. The difference of 157 basis points per year is usually referred to as survivorship bias.

To avoid this bias, many researchers dealing with mutual funds empirically use the survivorship bias-free database provided by the Center for Research in Security Prices (CRSP). Despite the existence of this database there are several reasons why it is important to analyze survivorship bias and the economic relations causing it on a broad and comprehensive basis. Firstly, virtually all pre-CRSP fund performance studies and a number of post-CRSP fund performance studies use survivorship biased data. To be able to compare the results to current studies and to interpret the results correctly, we have to learn more about the nature of survivorship bias, the differences resulting from different definitions and calculation methods, and the characteristics of survivors and non-survivors. Secondly, the survivorship bias itself reflects further economic insight, like the relation of fund size, survival, and performance which our study additionally analyzes in detail. Thirdly, there are mutual fund studies deliberately using survivorship biased data because of the specific methods used (e.g., Coles et al., 2006; Hübner, 2007; Bodson et al., 2008; Bessler et al., 2009). For these studies our paper provides guidance on how to reduce survivorship bias and how to interpret the results. Fourthly, the CRSP database only provides survivorship bias-free data for the US mutual fund market, whereas for many other countries such comprehensive data is not available (e.g., Gottesmann and Morey, 2007). Our study provides guidance on how to deal with the data selection problem. Finally, the survivorship problem does not exist exclusively in the mutual fund research area, but potentially affects all studies dealing with groups and portfolios of assets or other aggregated data for which survivorship bias-free data is not available.[1] Here, our study gives an idea which economic relations influence the scale of survivorship bias.

Because of its relevance, many studies on fund performance address the problem of survivorship bias. Besides a few papers analytically modeling the nature of survivorship bias (e.g., Berk and Green, 2004), the majority of related literature only deals with survivorship bias from an empirical perspective. Notably, there are very few studies systematically testing the significance of survivorship bias on comprehensive real fund return data.[2] Moreover, survivorship bias estimates, reported by former studies, range widely from 1 to 271 basis points per year (e.g., Grinblatt and Titman, 1989; Deaves, 2004).

Apart from different datasets used and different time periods covered, most studies define or calculate survivorship bias differently, making it difficult to compare

[1] Apart from studies on mutual fund performance, the problem also arises with respect to other financial instruments such as stocks (e.g., Brown et al., 1995; Boynton and Oppenheimer, 2006) or hedge funds (e.g., Brown et al., 1999; Liang, 2000; Malkiel and Saha, 2005; ter Horst and Verbeek, 2007; Eling, 2009).

[2] Grinblatt and Titman (1989) construct “hypothetical returns” based on quarterly fund holdings of a small fund sample. Carhart et al. (2002) only test the significance for one of their measures.

the results. Throughout the literature we do not find even two studies sharing identical empirical designs in determining survivorship bias. Therefore, the main focus of our paper is a comparison of different methodical approaches: (i) competing definitions of "survivor" and (ii) different weighting schemes for calculating fund portfolio returns. Besides this, the related literature shows alternative ways of aggregating individual funds and even different definitions of survivorship bias itself. Our contribution is the first study systematically calculating, testing, and explaining different definitions of the survivorship bias in fund performance.

In our empirical analysis, we examine the definitions and show the calculation methods applied as crucial for the magnitude of survivorship bias and the resulting differences regarded as significant. Moreover, we analyze the relation between fund size, survival, and performance to assess why different weighting schemes yield different survivorship bias estimates. We find the relation between survival and performance to be much stronger for small funds than for large funds. In addition, we analyze in detail the performance of non-survivors as the main driver of survivorship bias. We observe that non-survivors significantly under-perform and shrink in the years before disappearance. We also look at the performance of new funds to analyze a possible impact of incubation or backfill bias on our results. In doing so, we find some evidence for out-performance in new fund returns, but its impact on our findings is rather small and depends on the calculation method applied.

The remainder of this paper is organized as follows: Section 2 presents the definitions and methods commonly used in related literature and their implications for our empirical analysis. In Section 3 we describe the fund sample and report summary statistics. Section 4 presents and interprets our empirical results on survivorship bias, on the relation of fund size, survival, and performance, and on the performance of non-survivors and new funds. Section 5 summarizes and concludes.

2. Definitions and Methodology

2.1 DIFFERENT APPROACHES TO SURVIVORSHIP BIAS

Following the majority of previous studies, we define survivorship bias as the performance difference between two fund portfolios, a biased and an unbiased one. An unbiased portfolio consists of all funds operating at any time during the sample period (e.g., Elton et al., 1996; Blake and Timmermann, 1998; Carhart et al., 2002; ter Horst and Verbeek, 2007). This definition is appropriate for evaluating the historical performance of a portfolio including all funds investors were able to invest in over time.[3]

[3] According to this definition of an unbiased portfolio, a portfolio that does not include new funds (e.g., Grinblatt and Titman, 1989; Elton et al., 1996) is not unbiased. Also, the performance difference

A biased portfolio is a subset of the unbiased portfolio including only survivors which are defined in different ways in the literature. The first predominant approach is commonly known as end-of-sample conditioning. It defines all funds existing at the end of a specific sample period as survivors (e.g., Carhart et al., 2002). Wermers (1997), Blake and Timmermann (1998), ter Horst et al. (2001), Otten and Bams (2004), and Deaves (2004) also follow this approach. The second common survivor definition denotes only funds which operated through the whole sample period as survivors, henceforth "full-data conditioning" (e.g., Blake et al., 1993).[4] The full-data definition is used in studies by, e.g., Grinblatt and Titman (1989), Brown and Goetzmann (1995), Elton et al. (1996) and Holmes and Faff (2004). Malkiel (1995) uses both definitions.

The second main methodical difference in the literature is the weighting scheme for aggregating individual fund returns in a fund portfolio. The two commonly applied schemes are equal-weighting and value-weighting of individual fund returns by beginning of month total net assets (TNA). Despite many studies showing non-survivors to be smaller than survivors (e.g., Carhart, 1997; Zhao, 2005), most studies only use equal-weighted fund returns. Value-weighted fund returns are used in studies by, e.g., Brown and Goetzmann (1995), Malkiel (1995) and Deaves (2004).

To show the survivorship bias differences resulting from different methodologies, we estimate performance measures for six portfolios: the unbiased sample, the end-of-sample survivors, and the full-data survivors, where for each group an equal-weighted and a value-weighted portfolio are considered. The measures are based on time series representing the aggregate monthly returns of all funds allocated to the respective fund group (e.g., Carhart, 1997; Wermers, 1997; Carhart et al., 2002). This aggregation method allows us to use data on all funds regardless of the length of their return histories. Another advantage is that the time series of all portfolios have the same length and cover the same time period. Moreover, it allows us to directly use the monthly TNA of funds as weighting factors.

Another popular aggregation method not followed here is computing performance measures for all individual funds before averaging the performance of funds in the respective fund group (e.g., Elton et al., 1996; Carhart et al., 2002; Deaves, 2004). However, this approach is disadvantageous because it requires funds to have

between survivors and non-survivors examined as survivorship bias by Malkiel (1995) and Deaves (2004) and denoted as survivor premium by Blake and Timmermann (1998) does not match this definition because it does not describe the bias caused by ignoring non-survivors in a portfolio of funds.

[4] Another type of survivor definition is look-ahead conditioning (e.g., ter Horst et al., 2001; Carhart et al., 2002; ter Horst and Verbeek, 2007). To our knowledge, look-ahead bias plays a more important role in analyses of performance persistence as it requires funds to survive through two subsequent time periods. In a sense, look-ahead conditioning is a two period extension of full-data conditioning.

a return history of a certain length to generate reliable regression estimates. Funds not meeting this criterion, especially funds surviving only for a short period of time, are excluded systematically. Moreover, since individual funds partly exist in different time periods, their performance measures, in particular the mean excess return but also the Jensen alpha and the Fama/French alpha, might show a market climate bias (e.g., Scholz and Schnusenberg, 2009).[5]

2.2 PERFORMANCE MODELS

To show the performance of fund portfolios, we consider the majority of studies on survivorship bias and present results of four commonly used performance measures: The mean excess return MER_i with ER_{it} representing the total return net of expenses of a fund portfolio i in excess of the risk-free rate of return in month t.

$$MER_i = \frac{1}{T}\sum_{t=1}^{T} ER_{it} \tag{1}$$

The alpha from Jensen's (1968) one-factor model, with α_i representing the selection performance of fund portfolio i and ER_{mt} representing the market index return in excess of the risk-free rate of return in month t.

$$ER_{it} = \alpha_i + \beta_i ER_{mt} + \varepsilon_{it} \tag{2}$$

The alpha from the Fama and French (1993) three-factor model, where the two additional factors are the size factor *SMB* and the book-to-market factor *HML*.

$$ER_{it} = \alpha_i + \beta_{1i} ER_{mt} + \beta_{2i} SMB_t + \beta_{3i} HML_t + \varepsilon_{it} \tag{3}$$

The alpha from the Carhart (1997) four-factor model which includes the momentum factor *MOM* in addition to the three Fama and French factors.

$$ER_{it} = \alpha_i + \beta_{1i} ER_{mt} + \beta_{2i} SMB_t + \beta_{3i} HML_t + \beta_{4i} MOM_t + \varepsilon_{it} \tag{4}$$

2.3 ECONOMIC RELATIONS BEHIND SURVIVORSHIP BIAS

To obtain further economic insight into the nature of survivorship bias, we also analyze the relation between fund size, survival, and performance in detail. With respect to the relation between fund size and performance, related literature mainly discusses economies of scale (e.g., Indro et al., 1999) as a rationale for a positive relation, and liquidity disadvantages or ownership costs (e.g., Chen et al., 2004;

[5] Carhart et al. (2002) use both aggregation methods and obtain differing results applying Carhart's four-factor model. For the first approach (portfolio time series) and for the second approach (individual funds) they report annualized survivorship bias estimates of 96 basis points and 133 basis points, respectively. Therefore, the first approach can also be considered as more tentative.

Pollet and Wilson, 2008) as a rationale for a negative relation. The empirical evidence on the relation between fund size and performance is mixed. In favor of a negative relation Grinblatt and Titman (1989), Ciccotello and Grant (1996), and Chen et al. (2004) find that smaller US equity funds tend to out-perform larger funds explaining this relation with liquidity disadvantages.[6] Cremers and Petajisto (2009) who find a similar pattern explain their results with smaller funds being more active while larger funds tend towards indexing strategies.[7] On the other hand, Otten and Bams (2002) find the performance of European equity funds to be positively related to size. Indro et al. (1999) find mixed evidence and show that there is a positive size-performance relation up to an optimal fund size beyond which the relation becomes negative. Bird et al. (1983) also show mixed results for Australian funds. Grinblatt and Titman (1994) and Droms and Walker (1996) do not find significant relations at all.

In a fashion analog to Chen et al. (2004), we analyze the performance of equal-weighted and value-weighted size-decile portfolios which are created by resorting funds monthly to the respective portfolios according to their beginning of month TNA-ranks.[8] Based on aggregate monthly excess return time series of these size-decile portfolios, we find fund size to be in general positively related to performance.

With respect to the relation between size, performance, and survival, Brown and Goetzmann (1995) estimate probit models of fund disappearance as a function of specified variables such as returns, flows, size, expense ratios, and age. Using a pooled set of yearly fund data they find a positive relation between survival and returns as well as a positive relation between survival and size. In a first step, we reproduce this analysis based on a pooled set of yearly fund data derived from our total fund sample. We obtain very similar findings in terms of economic interpretation but also with respect to the actual figures. In a second step, we use a set of alternative model specifications which include dummy variables for large funds (upper third with respect to size in a given month) and small funds (lower third) to examine whether the relation between returns and survival is different for funds of different sizes. As a result, we find that returns have significantly higher impact on the survival of small funds than on the survival of large funds.

[6] Dahlquist et al. (2000) find that Swedish equity funds (Equity II) show a robust and strong negative size-performance relation. However, this could be due to the nature of the funds as “Equity II” stands for publicly managed funds. For common equity funds (Equity I) they find no significant effect.

[7] Cremers and Petajisto (2009) develop a new measure of active management, Active Share. They find Active Share to be positively related to performance and large funds to have a significantly smaller Active Share than smaller funds.

[8] As a robustness check, we also replicate the quintile-analysis of Chen et al. (2004). Our results on size-quintiles and size-deciles are economically the same. As deciles allow more detailed insight into the performance of very small and very large funds, we only report results on size-deciles. Size-quintile results are available on request.

Another important aspect for a better understanding of survivorship bias is the performance of non-survivors. Blake and Timmermann (1998) analyze return patterns over time for all disappeared UK equity funds showing that returns generally decrease towards disappearance. Zhao (2005) separately analyzes the performance of liquidated and non-liquidated US equity funds finding that liquidated funds under-perform non-liquidated funds. In our analysis, we combine both of these approaches and analyze the performance of all non-survivors, liquidated, and non-liquidated funds separately. For each of these three groups we create different time series including only fund returns of certain time frames before fund disappearance. By doing so these time series represent, for instance, only last-year or second-to-last-year fund returns, which allows us to show performance patterns over time.[9] These time series are compared to the time series of the end-of-sample survivors to determine the performance difference of non-survivors and survivors. We discover that (i) non-survivors significantly under-perform survivors in the years before disappearance, (ii) size and performance generally decrease towards disappearance, and (iii) liquidated funds significantly under-perform non-liquidated funds only in the last year before disappearance but not in the years before.

The difference between the alternative survivor definitions is mainly the exclusion of non-full-data survivors from the full-data survivor group. Therefore, we have to analyze in detail the performance of new funds to understand the survivorship bias differences from different definitions. Also, an out-performance of new funds could be an indicator for the existence of an incubation bias in our dataset (e.g., Evans, 2010). To evaluate the performance of new funds, we follow an approach similar to that for the performance of non-survivors and show performance patterns over time. The time series used represent, for example, first-year or second-year fund returns. These are compared to the time series of the complementary group of funds already existing at the beginning of the sample period, henceforth called "initial funds". We find that new funds show no significant out-performance.

3. Data and Summary Statistics

3.1 DATA

The market excess returns, the risk factors SMB, HML, and MOM, and the returns on the one-month US Treasury bill are provided by Kenneth R. French via an

[9] In their analysis of performance patterns over time Blake and Timmermann (1998) use a methodology from event study literature. We use a different approach to construct all time series used in our empirical study in an identical fashion. The resulting mean excess returns and alphas of these portfolios are directly comparable to the respective measures of other fund portfolios considered here. Moreover, our approach is based on fund portfolios an investor would have been able to invest in over time. Nevertheless, economic results in Blake and Timmermann (1998) are similar to ours.

online data library.[10] As source of fund data we use the CRSP Survivor-Bias-Free US Mutual Fund Database. Our evaluation period starts in January 1993 and ends in December 2006. We extract the analyzed sample applying the following selection criteria. We exclude funds not continuously classified as US domestic equity by the Standard & Poors fund objective code.[11] We eliminate funds with fragmentary return histories or unreliable returns[12] as well as funds without any available TNA data. Funds with multiple share classes are merged by portfolio codes. Our final sample contains 4,964 US domestic equity mutual funds.

A problem we face with monthly TNA is that this data is incomplete for about one-third of the funds in our sample. We therefore fill in missing values in order to have a complete set of weighting factors. To do this, we follow a three-step procedure. Firstly, we compute monthly value-weighted average fund growth rates for the period 01/1993 through 12/2006 based on the given data. Secondly, we extrapolate values missing at the beginning and at the end of the TNA time series of individual funds, applying these average fund growth rates. Thirdly, we fill gaps within individual fund time series by geometric interpolation, assuming constant relative growth between known data points. In total, we fill less than 3.3% of the monthly values or less than 1.3% of TNA (in US$), respectively. For less than 2% of all funds, we fill more than 24 months of missing values.

Furthermore, we observe seemingly implausible jumps in some of the TNA time series. 40 funds show TNA jumping to more than ten times their size and back in the following month, or vice versa. This may be due to reporting or recording errors. However, since these jumps have no impact on equal-weighted results we do not exclude these funds from our original analysis.

As robustness checks, we replicate our analysis on survivorship bias and survivorship bias differences (i) without filling missing values, (ii) including filled values but excluding funds with more than 24 monthly values filled, (iii) including filled values but excluding funds with jumps in TNA, and (iv) without filling values combined with excluding funds with more than 24 missing values and excluding funds with jumps in TNA. Apart from very small changes in the actual figures of the four robustness checks, the economics remain unchanged.[13]

---

[10] The data library is available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

[11] The Standard & Poors US domestic equity mutual fund objective codes are Equity USA Aggressive Growth (AGG), Equity USA Midcap (GMC), Equity USA Growth & Income (GRI), Equity USA Growth (GRO), Equity USA Income & Growth (ING), Equity USA Small Companies (SCG) and Asset allocation USA Preferred (CPF).

[12] Monthly returns higher (lower) than 50% (−50%) are, if possible, checked for plausibility by comparison with Morningstar data. Funds are erased if (1) suspicion is confirmed by Morningstar data or (2) the suspicious data does not fit into the overall return history of the fund or the cross-sectional return distribution of the month.

[13] The results of the robustness checks are available on request.

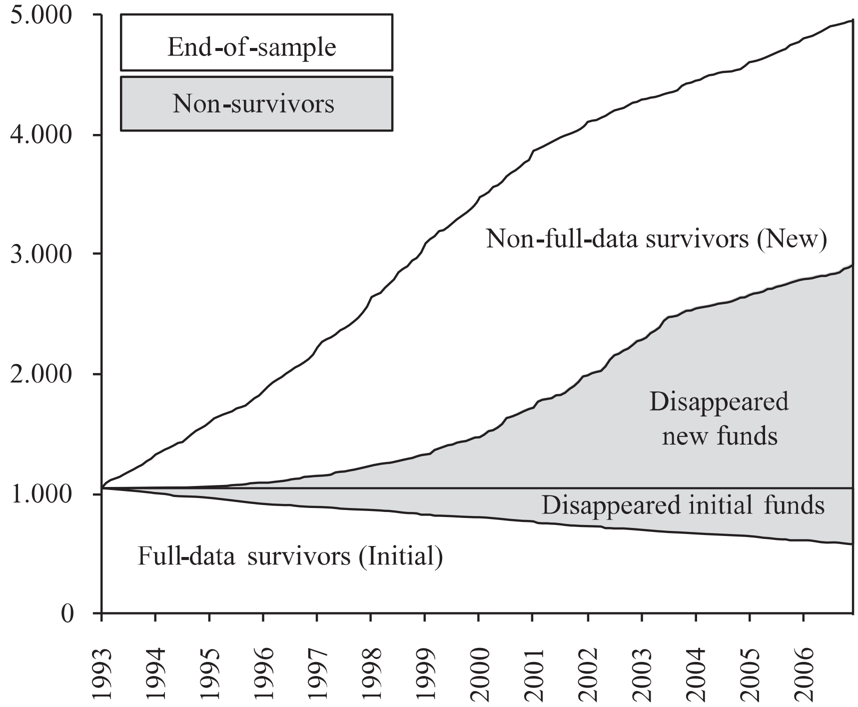

Figure 1. Number of funds in different fund groups over time.
This figure shows the number of funds allocated to four different fund groups and the development of the fund sample over time in the period from 01/1993 through 12/2006. In total, the sample consists of 4,964 funds. In 01/1993 the initial sample includes 1,054 funds of which 597 survive (full-data survivors) and 457 disappear. After 01/1993, 3,910 new funds enter the sample of which 2,044 survive (non-full-data survivors) and 1,866 disappear. Full-data and non-full-data survivors together add up to 2,641 end-of-sample survivors. Disappeared initial funds and disappeared new funds together add up to 2,323 non-survivors.

3.2 SUMMARY STATISTICS

Figure 1 reports the total number of funds in our overall sample over time from 01/1993 through 12/2006. The figure also shows how this unbiased sample of 4,964 funds divides into different fund groups. In 01/1993 the sample starts with 1,054 initial funds (390 billion US$). Of these, 597 funds survive showing full data (full-data survivors) and 457 disappear (disappeared initial funds). After 01/1993, the US domestic equity mutual fund market substantially grows with 3,910 new funds entering the sample. Of these, 2,044 survive (non-full-data survivors) and 1,866 disappear (disappeared new funds). Together, full-data survivors and non-full-data survivors add up to 2,641 end-of-sample survivors (3,476 billion US$) in 12/2006. Disappeared initial funds and disappeared new funds add up to 2,323 non-survivors.

Table I shows sample summary statistics on the fund groups shown in Figure 1. Focusing on the first four groups, which are most important with respect to our survivorship bias analysis, shows that during the covered sample period the average (median) fund in the unbiased sample operates for 78 (67) months, end-of-sample survivors for 98 (99) months, and non-survivors for 54 (45) months. By definition,

Table I. Sample summary statistics

This table shows summary statistics for the total fund sample and for the subdivisions shown in Figure 1 in the evaluation period from 01/1993 through 12/2006. Excess returns are denoted in percentage points per month. Monthly mean excess returns are calculated based on the aggregated monthly excess return time series of a respective equal-weighted fund portfolio. The medians and standard deviations are calculated analogously. TNA are denoted in million US$. Mean TNA are calculated based on the aggregated mean TNA time series of a respective equal-weighted fund portfolio. The medians and standard deviations are calculated analogously.

	Unbiased sample	End-of-sample survivors	Full-data survivors	Non-full-data survivors	Non-survivors	New funds	Initial funds	Initial disappeared funds	New disappeared funds
Number of funds	4,964	2,641	597	2,044	2,323	3,910	1,054	457	1,866
Relative number of funds (%)	100.0	53.2	12.0	41.2	46.8	78.8	21.2	9.2	37.6
Monthly return observations	385,592	259,771	100,296	159,475	125,821	248,874	136,718	36,422	89,399
Average fund life in sample (months)	78	98	168	78	54	64	130	80	48
Median fund life in sample (months)	67	99	168	78	45	56	168	78	40
Monthly excess return (%)									
Mean	0.5375	0.6577	0.5994	0.7155	0.2859	0.5477	0.5330	0.2222	0.3099
Median	1.0906	1.0973	1.1593	1.2402	0.7625	1.1329	1.1047	0.7447	0.9430
Standard deviation	4.1501	4.0943	4.0034	4.1675	4.2436	4.2090	4.0353	4.1623	4.2763
TNA (*million US$*)									
Absolute in 12/2006	3,475,875	3,475,875	2,425,353	1,050,523	–	1,050,523	2,425,353	–	–
Relative in 12/2006 (%)	100.0	100.0	69.8	30.2	–	30.2	69.8	–	–
Mean	775	1,079	2,407	274	183	203	2,042	565	68
Median	795	1,135	2,628	280	176	193	2,241	599	62
Standard deviation	260	293	1,061	127	44	128	1,037	265	34

all full-data survivors exist through the entire period of 168 months. Of the other fund groups, new disappeared funds show the shortest fund life with only 48 (40) months of existence during the evaluation period.

Regarding monthly excess returns, we find distinct differences between the first four fund groups. End-of-sample survivors (esp. non-full-data survivors) show the highest returns while non-survivors (esp. initial disappeared funds) show the lowest. At the same time, non-survivors (esp. new disappeared funds) show the highest standard deviation.

In terms of size, there are even more pronounced differences between the fund groups. Non-survivors hold mean (median) TNA of 183 (176) million US$ while operating, end-of-sample survivors hold 1,079 (1,135) million US$, and full-data survivors hold even more with 2,407 (2,628) million US$. In relative numbers, full-data survivors represent only 12% of the total number of funds but account for more than two thirds (69.8%) of the fund market's TNA in 12/2006. Overall, the smallest funds by far are new disappeared funds followed by new funds in general. Due to these distinct differences between the fund groups in terms of size and performance, we also expect different survivorship bias estimates depending on survivor definitions and weighting schemes.

Table II reports annual fund starts and fund disappearances through our sample period from 01/1993 through 12/2006 as well as average numbers for the whole sample period and sub-periods.[14] A comparison of the two seven-year sub-periods reveals that relative annual fund starts decreased by more than half from 20.03% to 8.03%. On the other hand, relative annual fund disappearances grew from 5.25% to 8.64%. For the period from 1962 through 1995 Carhart (1997) reports an even smaller annual fund attrition rate of 3.6%. Eling (2008) also reports an increase in attrition rates for stock and bond funds from 1996 to 2005. Overall, it seems like shutting funds is becoming a more common practice among fund companies over time. Thus, for future performance studies survivorship bias should become more important.

4. Empirical Results

4.1 SURVIVORSHIP BIAS

Table III reports performance measures for seven differently composed fund groups. Those are the unbiased sample, the end-of-sample survivors and the full-data survivors, as well as the non-full-data survivors, non-survivors, new funds, and initial funds. For each group an equal-weighted portfolio (Panel I) and a value-weighted

[14] Note that "fund start" stands for "fund history starts in CRSP", and "fund disappearance" stands for "fund history ends in CRSP" (e.g., Amin and Kat, 2003).

Table II. Annual fund starts and fund disappearances

This table shows the absolute and relative numbers of annual fund starts and fund disappearances through the sample period from 01/1993 through 12/2006. Relative numbers refer to the total number of funds operating in December of the prior year. Average numbers refer to the mean of the respective annual numbers.

	Fund starts		Fund disappearances	
Year	Absolute	Relative	Absolute	Relative
1993	247	23.43%	46	4.36%
1994	276	21.89%	47	3.73%
1995	252	16.96%	83	5.59%
1996	343	20.68%	82	4.94%
1997	400	20.87%	109	5.69%
1998	454	20.49%	131	5.91%
1999	402	15.87%	166	6.55%
2000	372	13.44%	279	10.08%
2001	275	9.59%	309	10.77%
2002	211	7.45%	330	11.66%
2003	163	6.00%	290	10.67%
2004	137	5.28%	135	5.21%
2005	202	7.76%	167	6.42%
2006	176	6.71%	149	5.68%
Average 1993–2006	279	14.03%	166	6.95%
Sub-average 1993–1999	339	20.03%	95	5.25%
Sub-average 2000–2006	219	8.03%	237	8.64%

portfolio (Panel II) are considered. Among the first three groups, which are most relevant in our survivorship bias analysis, the end-of-sample survivors show the highest mean excess return. The mean excess returns of the unbiased portfolios (equal-weighted and value-weighted) are clearly below those of the respective biased portfolios. In terms of risk-adjusted performance almost all unbiased alphas are negative and significantly different from zero; except for the equal-weighted Jensen alpha which is also negative but insignificant. In contrast, the equal-weighted end-of-sample survivor portfolio shows a positive Jensen alpha. This means that besides performance differences there is also potential for misinterpreting the average fund as having out-performed the passive benchmark during the sample period.

For the unbiased sample, value-weighting shows higher performance than equal-weighting. This is because the poor performance of the small non-survivors is overemphasized when equal-weighted. For the biased group of end-of-sample survivors equal-weighting shows higher performance than value-weighting. This might be because poorly performing small funds drop out of the sample and well performing large funds are underemphasized with equal-weighting. For full-data survivors there is no clear domination of one weighting scheme, which might be due to full-data survivors being a more homogeneous group in terms of fund size.

Table III. Fund portfolio performance

This table shows equal-weighted (Panel I) and value-weighted (Panel II) performance measures and factor loadings for seven fund groups in the period from 02/1993 through 12/2006. The unbiased sample consists of all funds operating at any time during the sample period. Full-data and non-full-data survivors together add up to the end-of-sample survivors, which in combination with non-survivors add up to the unbiased sample. New funds and initial funds also add up to the unbiased sample. All results are based on monthly aggregated excess return time series. Value-weighted returns are weighted by beginning of month TNA. Performance measures are quoted in percentage points per month. p-values are reported in parentheses and computed using two-sided t-tests for means and two-sided t-tests for regression coefficients, respectively. p-values for regression coefficients are based on HAC-consistent covariances (Newey and West, 1987). In Panel I, adjusted R^2-statistics range from 0.94 (non-full-data survivors, 1F-Model) to 0.98 (several 3F- and 4F- results). In Panel II, adjusted R^2-statistics are between 0.95 (new funds, 1F-Model) and 0.99 (several 1F-, 3F-, and 4F-results).

		1F-Model		3F-Model				4F-Model				
	MER (%)	Alpha (%)	ER_m	Alpha (%)	ER_m	*SMB*	*HML*	Alpha (%)	ER_m	*SMB*	*HML*	*MOM*
Panel I. Equal-weighted												
Unbiased portfolio	0.5375	−0.0907	0.9879	−0.1556	1.0082	0.1239	0.0950	−0.1790	1.0141	0.1249	0.0952	0.0249
	(0.097)	(0.213)	(0.000)	(0.002)	(0.000)	(0.008)	(0.001)	(0.001)	(0.000)	(0.007)	(0.001)	(0.183)
End-of-sample survivors	0.6577	0.0395	0.9721	−0.0470	1.0020	0.1333	0.1208	−0.0778	1.0097	0.1346	0.1210	0.0328
	(0.040)	(0.628)	(0.000)	(0.383)	(0.000)	(0.005)	(0.000)	(0.140)	(0.000)	(0.004)	(0.000)	(0.010)
Full-data survivors	0.5994	−0.0093	0.9572	−0.0974	0.9906	0.1008	0.1164	−0.1073	0.9930	0.1012	0.1165	0.0104
	(0.055)	(0.896)	(0.000)	(0.025)	(0.000)	(0.003)	(0.000)	(0.024)	(0.000)	(0.003)	(0.000)	(0.503)
Non-full-data survivors	0.7155	0.0915	0.9812	0.0102	1.0065	0.1582	0.1196	−0.0336	1.0174	0.1600	0.1200	0.0467
	(0.028)	(0.325)	(0.000)	(0.879)	(0.000)	(0.007)	(0.001)	(0.594)	(0.000)	(0.005)	(0.001)	(0.060)
Non-survivors	0.2859	−0.3551	1.0083	−0.3737	1.0094	0.0934	0.0384	−0.3868	1.0126	0.0939	0.0385	0.0137
	(0.388)	(0.000)	(0.000)	(0.000)	(0.000)	(0.044)	(0.186)	(0.000)	(0.000)	(0.042)	(0.201)	(0.463)
New funds	0.5477	−0.0864	0.9970	−0.1468	1.0136	0.1430	0.0937	−0.1751	1.0207	0.1442	0.0939	0.0302
	(0.096)	(0.276)	(0.000)	(0.012)	(0.000)	(0.008)	(0.002)	(0.003)	(0.000)	(0.006)	(0.003)	(0.154)
Initial funds	0.5330	−0.0812	0.9660	−0.1552	0.9933	0.0925	0.0992	−0.1679	0.9965	0.0930	0.0993	0.0136
	(0.091)	(0.226)	(0.000)	(0.000)	(0.000)	(0.007)	(0.000)	(0.000)	(0.000)	(0.007)	(0.000)	(0.380)

(*continued*)

Table III. (Continued)

		1F-Model		3F-Model				4F-Model				
	MER (%)	Alpha (%)	ER_m	Alpha (%)	ER_m	*SMB*	*HML*	Alpha (%)	ER_m	*SMB*	*HML*	*MOM*
Panel II. Value-weighted												
Unbiased portfolio	0.5388	−0.0847	0.9805	−0.1044	0.9882	0.0194	0.0254	−0.1240	0.9931	0.0203	0.0255	0.0208
	(0.089)	(0.016)	(0.000)	(0.002)	(0.000)	(0.264)	(0.051)	(0.000)	(0.000)	(0.239)	(0.054)	(0.019)
End-of-sample survivors	0.5613	−0.0591	0.9755	−0.0861	0.9868	0.0184	0.0334	−0.1034	0.9912	0.0191	0.0335	0.0184
	(0.075)	(0.087)	(0.000)	(0.009)	(0.000)	(0.226)	(0.006)	(0.002)	(0.000)	(0.208)	(0.006)	(0.033)
Full-data survivors	0.5575	−0.0559	0.9645	−0.0809	0.9762	0.0025	0.0281	−0.0900	0.9785	0.0029	0.0282	0.0096
	(0.074)	(0.096)	(0.000)	(0.011)	(0.000)	(0.820)	(0.009)	(0.007)	(0.000)	(0.802)	(0.008)	(0.299)
Non-full-data survivors	0.5373	−0.1156	1.0237	−0.1347	1.0239	0.1005	0.0392	−0.1929	1.0388	0.1030	0.0403	0.0624
	(0.112)	(0.070)	(0.000)	(0.018)	(0.000)	(0.050)	(0.168)	(0.000)	(0.000)	(0.030)	(0.171)	(0.023)
Non-survivors	0.2926	−0.3592	1.0254	−0.3036	0.9977	0.0379	−0.0607	−0.3448	1.0080	0.0113	−0.0604	0.0440
	(0.383)	(0.000)	(0.000)	(0.000)	(0.000)	(0.801)	(0.015)	(0.000)	(0.000)	(0.756)	(0.025)	(0.004)
New funds	0.5028	−0.1557	1.0325	−0.1657	1.0278	0.1052	0.0296	−0.2245	1.0428	0.1077	0.0306	0.0631
	(0.141)	(0.023)	(0.000)	(0.005)	(0.000)	(0.052)	(0.316)	(0.000)	(0.000)	(0.032)	(0.330)	(0.028)
Initial funds	0.5388	−0.0769	0.9681	−0.0957	0.9769	0.0021	0.0213	−0.1066	0.9796	0.0026	0.0214	0.0116
	(0.085)	(0.021)	(0.000)	(0.003)	(0.000)	(0.868)	(0.068)	(0.001)	(0.000)	(0.845)	(0.066)	(0.209)

When equal-weighted, end-of-sample survivors perform better than full-data survivors. A first explanation could be the difference in fund sizes. As the full-data survivors majorly include large funds, while the end-of-sample survivors additionally include a large number of smaller funds, a survival criterion that eliminates small under-performing funds but keeps large under-performing funds alive could explain why end-of-sample survivors out-perform full-data survivors. Therefore, in a later section we analyze the relationship between size, survival, and performance in more detail.

Regarding the last four fund groups in Table III, the clearly above-average performance of the non-full-data survivors on an equal-weighted basis is striking. This result explains why end-of-sample survivors out-perform full-data survivors when equal-weighted. A rationale for this could be the practice of incubating new funds. Incubation means that fund families open new funds to the public only after a certain incubation period during which new funds internally create favorable return histories with very little money (e.g., Arteaga et al., 1998; Deaves, 2004; Malkiel and Saha, 2005; Evans, 2006; Evans, 2010; Karoui and Meier, 2009). When opened to the public these histories are backfilled into fund databases. In contrast, funds with inferior performance histories are not opened to the public after the incubation period, and no return histories are reported to fund databases. However, comparing the equal-weighted results to the value-weighted results shows that the above-average performance of non-full-data survivors almost completely vanishes, thereby explaining the small difference between end-of-sample and full-data survivors when value-weighting is considered. This pattern could be attributed to a "new-fund-survivorship bias", whereby predominantly out-performing but relatively small new funds (both incubated and regularly opened) survive. Comparing the results of new funds and initial funds reveals that there is some out-performance of new funds on an equal-weighted basis, though not as strong as before. Value-weighted initial funds even out-perform new funds. Therefore, we will further investigate a possible impact of incubation bias on our empirical results in a later section.

Comparing the results on non-survivors and non-full-data survivors on both equal-weighted and value-weighted bases could also be evidence for the presumption that smaller and younger funds take higher risk. Cremers and Petajisto (2009) support this by finding that smaller and younger funds are more actively managed. If this is the case, the performance distribution of smaller funds has fatter tails such that small survivors show extreme out-performance, which is another clue towards equal-weighted incubation or new-fund-survivorship bias. Analog to this small non-survivors show extreme under-performance which could be a rationale for equal-weighted survivorship bias.

Table III also shows factor loadings for the seven fund groups. Here, we concentrate on full-data survivors, non-full-data survivors, and non-survivors because together these three disjoint groups add to the unbiased sample. Moreover, we only

Table IV. Survivorship bias estimates

This table shows equal-weighted (Panel I) and value-weighted (Panel II) survivorship bias estimates for the period from 02/1993 through 12/2006. All survivorship bias estimates are based on the time series of differences between biased and unbiased portfolios (e.g., equal-weighted end-of-sample less equal-weighted unbiased). Value-weighted returns are weighted by the beginning of month TNA. Survivorship biases are quoted in percentage points per month. Reported p-values are computed using two-sided t-tests for means and two-sided t-tests for regression coefficients, respectively. p-values for regression coefficients are based on HAC-consistent covariances (Newey and West, 1987).

	End-of-sample survivorship		Full-data survivorship	
	Bias (%)	p-value	Bias (%)	p-value
Panel I. Equal-weighted				
MER	0.1202	0.000	0.0618	0.009
1F-Alpha	0.1302	0.000	0.0814	0.001
3F-Alpha	0.1086	0.000	0.0582	0.000
4F-Alpha	0.1012	0.000	0.0718	0.000
Panel II. Value-weighted				
MER	0.0255	0.000	0.0187	0.136
1F-Alpha	0.0257	0.000	0.0288	0.008
3F-Alpha	0.0183	0.000	0.0235	0.003
4F-Alpha	0.0206	0.000	0.0340	0.000

consider value-weighted portfolios as these represent the fund market more accurately. In general, non-full-data survivors and non-survivors show higher exposure to the market. Further, there are differences in the exposures to HML. Full-data survivors show positive and significant exposure to HML assuming an aggregate value strategy. Non-survivors on the other hand show a negative and significant exposure to HML assuming an aggregate growth strategy. Non-full-data survivors show no significant exposure to HML, but in contrast to the other fund groups they show a positive and significant exposure to SMB assuming an aggregate small-cap strategy. Also, full-data survivors show no significant momentum strategy while the non-full-data survivors and non-survivors show positive and significant loadings on MOM.

As our main results of this section, Table IV reports survivorship bias estimates. Across all performance measures and method combinations, survivorship bias is positive and, with just one exception, significantly different from zero. This confirms previous research that survivorship bias in general leads to overstating fund portfolio performance. Statistically significant results range from 22 to 157 basis points per year[15] (1.83 to 13.02 basis points per month). These numbers also confirm previous research, where the majority of survivorship bias estimates following a definition similar to ours range from 20 to 150 basis points per year.

[15] Annualized by $SB_{yearly} = (1 + SB_{monthly})^{12} - 1$ (e.g., Deaves, 2004).

Distinguishing different weighting-schemes, equal-weighting always yields higher survivorship bias estimates than value-weighting. With 70 to 157 basis points per year (5.82 to 13.02 basis points per month), the equal-weighted survivorship bias is both significantly different from zero and economically relevant. Value-weighted survivorship bias with 22 to 41 basis points per year (1.83 to 3.40 basis points per month) is still statistically significant, but seems of minor economic importance. Concerning method combinations, the equal-weighted end-of-sample survivorship biases are the highest with 122 to 157 basis points per year (10.12 to 13.02 basis points per month) depending on the measure used. The value-weighted end-of-sample estimates are lowest for three of four performance measures with 22 to 31 basis points per year (1.83 to 2.57 basis points per month).

Concerning different performance measures, there seem to be obvious differences in the survivorship bias estimates. For example, the Jensen alpha shows the highest end-of-sample survivorship bias (equal-weighted and value-weighted), the highest equal-weighted full-data survivorship bias, and the second-highest value-weighted full-data survivorship bias. This can be explained based on an approach used in Pástor and Stambaugh (2002).[16] In the following we only consider the equal-weighted end-of-sample portfolio (*ewEoS*) but the same principle applies to the other portfolios as well.

Assuming the Carhart model to describe the return generating process of the portfolio, its Jensen alpha (α^{1F}_{ewEoS}) can be decomposed as follows: first, the Carhart alpha (α^{4F}_{ewEoS}), and second, the four-factor loadings of the equal-weighted end-of-sample portfolio on the factors *SMB*, *HML* and *MOM* (e.g., β^{SMB-4F}_{ewEoS} for *SMB*) times the Jensen alphas of the respective factors themselves (e.g., α^{1F}_{SMB}). These factor alphas (displayed in Table V) are obtained by applying the Jensen model to the time series of the respective factor (e.g., *SMB*). They describe the relationship between the factors and the market index and reflect the general market environment in the sample period.

$$\alpha^{1F}_{ewEoS} = \alpha^{4F}_{ewEoS} + \beta^{SMB-4F}_{ewEoS}\alpha^{1F}_{SMB} + \beta^{HML-4F}_{ewEoS}\alpha^{1F}_{HML} + \beta^{MOM-4F}_{ewEoS}\alpha^{1F}_{MOM} \quad (5)$$

Following the same approach, the Fama/French alpha of the equal-weighted end-of-sample portfolio (α^{3F}_{ewEoS}) can be decomposed into the Carhart alpha (α^{4F}_{ewEoS}) and the four-factor loading on *MOM* (β^{MOM-4F}_{ewEoS}) times the Fama/French alpha of the factor *MOM* itself (α^{3F}_{MOM}).

$$\alpha^{3F}_{ewEoS} = \alpha^{4F}_{ewEoS} + \beta^{MOM-4F}_{ewEoS}\alpha^{3F}_{MOM} \quad (6)$$

Equations (5) and (6) clearly show that the differences between the one-, three-, and four-factor alphas depend on the four-factor loadings of the equal-weighted end-of-sample portfolio and the factor alphas. Table III shows positive four-factor

[16] See also, e.g., Matallín-Sáez (2006) and Scholz and Schnusenberg (2009).

Table V. Performance measures of risk factors

This table shows performance measures and factor loadings of risk factors in the period from 02/1993 through 12/2006. All results are based on monthly risk factor time series. Performance measures are calculated for the risk factors not included in the respective model and quoted in percentage points per month. p-values are reported in parentheses and computed using two-sided t-tests for means and two-sided t-tests for regression coefficients, respectively. p-values for regression coefficients are based on HAC-consistent covariances (Newey and West, 1987).

		1F-Model		3F-Model			
	MER (%)	Alpha (%)	ER_m	Alpha (%)	ER_m	*SMB*	*HML*
ER_m	0.6359						
	(0.048)						
SMB	−0.0638	−0.1690	0.1654				
	(0.832)	(0.571)	(0.008)				
HML	0.6293	0.9028	−0.4300				
	(0.021)	(0.003)	(0.000)				
MOM	0.7868	0.9376	−0.2372	0.9381	−0.2341	−0.0395	−0.0080
	(0.044)	(0.004)	(0.157)	(0.021)	(0.110)	(0.804)	(0.971)

loadings of the upper three portfolios on all factors. Table V shows non-zero alphas for the time series of the factors *SMB*, *HML* and *MOM*. Together, this leads to different alphas from different models.

The decomposition approach underlying Equations (5) and (6) can be transferred from alpha to survivorship bias. In the case of equal-weighted end-of-sample one-factor survivorship, the bias is defined as the difference between the Jensen alpha of the equal-weighted end-of-sample portfolio and the Jensen alpha of the equal-weighted unbiased portfolio ($SB^{1F}_{ewEoS} = \alpha^{1F}_{ewEoS} - \alpha^{1F}_{unbiased}$). Comparable to Equation (5), the one-factor survivorship bias can be presented as the four-factor survivorship bias (SB^{4F}_{ewEoS}) plus the differences of the four-factor loadings of biased and unbiased portfolios (e.g., $\beta^{SMB-4F}_{ewEoS} - \beta^{SMB-4F}_{unbiased}$ for *SMB*) times the Jensen alphas of the factors (e.g., α^{1F}_{SMB}).

$$\begin{aligned} SB^{1F}_{ewEoS} = SB^{4F}_{ewEoS} &+ \left(\beta^{SMB-4F}_{ewEoS} - \beta^{SMB-4F}_{unbiased}\right)\alpha^{1F}_{SMB} + \left(\beta^{HML-4F}_{ewEoS} - \beta^{HML-4F}_{unbiased}\right)\alpha^{1F}_{HML} \\ &+ \left(\beta^{MOM-4F}_{ewEoS} - \beta^{MOM-4F}_{unbiased}\right)\alpha^{1F}_{MOM} \end{aligned} \tag{7}$$

Analog to this, the three-factor survivorship bias (SB^{3F}_{ewEoS}) can be decomposed as in Equation (6) into the four-factor survivorship bias and the difference between the factor loadings of the biased and the unbiased portfolio on *MOM* ($\beta^{MOM-4F}_{ewEoS} - \beta^{MOM-4F}_{unbiased}$) times the Fama/French alpha of the factor *MOM* (α^{3F}_{MOM}).

$$SB^{3F}_{ewEoS} = SB^{4F}_{ewEoS} + \left(\beta^{MOM-4F}_{ewEoS} - \beta^{MOM-4F}_{unbiased}\right)\alpha^{3F}_{MOM} \tag{8}$$

Equations (7) and (8) clearly show that the differences between survivorship biases from different performance models depend on the differences in the four-factor

Table VI. Differences in survivorship bias estimates

This table shows survivorship bias differences in the period from 02/1993 through 12/2006. Panel I reports survivorship bias differences resulting from different survivor definitions. The differences are determined by the time series of differences between the end-of-sample portfolio and the full-data portfolio, each equal-weighted and value-weighted. Panel II shows survivorship bias differences resulting from different weighting schemes. We determine these differences based on the time series of differences between the differential time series displayed in Table IV (e.g., (equal-weighted full-data less equal-weighted unbiased) less (value-weighted full-data less value-weighted unbiased)). Value-weighted returns are weighted by the beginning of month TNA. Survivorship bias differences are quoted in percentage points per month. Reported p-values are computed using two-sided t-tests for means and two-sided t-tests for regression coefficients, respectively. p-values for regression coefficients are based on HAC-consistent covariances (Newey and West, 1987).

Panel I. End-of-sample less full-data

	Equal-weighted		Value-weighted	
	Difference (%)	p-value	Difference (%)	p-value
MER	0.0584	0.006	0.0038	0.687
1F-Alpha	0.0489	0.001	−0.0032	0.657
3F-Alpha	0.0504	0.007	−0.0052	0.388
4F-Alpha	0.0295	0.050	−0.0135	0.020

Panel II. Equal-weighted less value-weighted

	End-of-sample		Full-data	
	Difference (%)	p-value	Difference (%)	p-value
MER	0.0977	0.000	0.0432	0.001
1F-Alpha	0.1046	0.000	0.0525	0.002
3F-Alpha	0.0903	0.000	0.0347	0.003
4F-Alpha	0.0807	0.000	0.0377	0.001

loadings of the biased and the unbiased portfolio as well as the respective factor alphas which reflect the market environment during the sample period.

The results in Table III show that there are differences between the fund groups such that the unbiased sample loads higher on the market excess return than both survivor groups. On the other hand, the unbiased sample has the lowest loadings on *HML* compared to the survivor groups. The full-data survivors show the lowest loadings on *SMB* and on *MOM* and the end-of-sample survivors exhibit the highest loadings on *HML*. As stated before, the different survivor groups exhibit different investment styles.

Table VI analyzes and tests the differences between the combinations of methods by reporting differences between end-of-sample and full-data survivorship biases (Panel I) as well as differences between equal-weighted and value-weighted results (Panel II). Panel I shows that the end-of-sample survivorship bias is significantly higher than the full-data survivorship bias for equal-weighting. For value-weighting there is no clear domination of end-of-sample or full-data survivorship. The results are of mixed signs, rather small and for three of four performance measures statistically not different from zero. Only for the four-factor model do we find a significant but still rather small difference. This could be because using value-weighted returns, the disproportional impact of out-performing but small new funds (both incubated and regularly opened) does not occur.

Panel II of Table VI presents more distinct relations because for both survivor definitions the equal-weighted survivorship bias estimates are substantially higher than the value-weighted results. This is not surprising, as under-performing small non-survivors have a stronger impact on the performance of the unbiased sample in the case of equal-weighting. All differences are significantly different from zero regardless of survivor definition and performance measure. In the case of end-of-sample conditioning, equal-weighted survivorship biases exceed value-weighted estimates by 97 to 126 basis points per year (8.07 to 10.46 basis points per month). This means that the equal-weighted survivorship bias is approximately five times higher, which clearly is of economic interest. With full-data conditioning, equal-weighted survivorship biases exceed value-weighted estimates by 42 to 63 basis points per year (3.47 to 5.25 basis points per month), which is still more than twice the size and therefore also economically relevant.

4.2 FUND SIZE, SURVIVAL, AND PERFORMANCE

As there are systematic differences between survivorship bias estimates from different weighting schemes and from different survivor definitions, the relation between fund size, survival, and performance is worth a detailed analysis. Table VII presents mean TNA as well as the mean excess return, the Jensen alpha, the Fama/French alpha, and the Carhart alpha for size-decile portfolios. We find that the smallest 10% of funds perform worst for all performance measures and weighting schemes.[17] Moreover, we find that the upper middle-size-deciles perform best in cases of the mean excess return and the Jensen alpha. For the Fama/French alpha and the Carhart alpha the largest 10% of funds out-perform most other deciles.

The last three columns of Table VII give new insight into the disappearance of funds. The results describe fund disappearance rates expressed by the proportion of funds allocated to a certain size-decile at any point in time that disappear within a one-month ($t + 1m$), one-year ($t + 1y$), or two-year period afterwards ($t + 2y$), respectively. The numbers clearly show that the fund disappearance rates increase

[17] We test the performance differences between the smallest decile and the other nine deciles and reject the H0 that the difference is zero in all cases. Also, we test the performance difference between the largest decile and the other deciles (except the smallest). Here, all differences are not significantly different from zero. All test results are available on request.

Table VII. Performance of size-decile portfolios

This table shows statistics on size-decile portfolios in the period from 02/1993 through 12/2006. These portfolios are created by rebalancing individual funds monthly on the basis of their beginning of month TNA-ranks. The first decile represents the largest 10% of funds; the tenth decile represents the smallest 10%. Mean TNA (MTNA) are denoted in million US$ and represent the mean of the aggregated TNA time series of a respective size-decile portfolio. All performance measures are based on monthly aggregate excess return time series. Value-weighted returns are weighted by the beginning of month TNA. p-values are reported in parentheses and computed using two-sided t-tests for means and two-sided t-test for regression coefficients, respectively. p-values for regression coefficients are based on HAC-consistent covariances (Newey and West, 1987). The last three columns show the disappearance rate of funds allocated to a certain size-decile at any point in time within a one-month ($t+1m$), a one-year ($t+1y$), or a two-year period afterwards ($t+2y$), respectively.

		Equal-weighted performance (%)				Value-weighted performance (%)				Disappearance rates (%)		
Size-decile	MTNA (Mio US$)	MER	1F-Alpha	3F-Alpha	4F-Alpha	MER	1F-Alpha	3F-Alpha	4F-Alpha	$(t+1m)$	$(t+1y)$	$(t+2y)$
1 (Large)	5,968	0.5325	−0.0962	−0.1196	−0.1399	0.5329	−0.0781	−0.0901	−0.1025	0.01	0.26	0.64
		(0.094)	(0.015)	(0.001)	(0.000)	(0.088)	(0.019)	(0.006)	(0.002)			
2	879	0.5500	−0.1039	−0.1452	−0.2004	0.5495	−0.1032	−0.1446	−0.1986	0.07	0.95	2.26
		(0.101)	(0.140)	(0.009)	(0.000)	(0.102)	(0.138)	(0.009)	(0.000)			
3	410	0.5388	−0.0969	−0.1605	−0.2055	0.5392	−0.0965	−0.1582	−0.2015	0.17	2.26	5.22
		(0.099)	(0.172)	(0.002)	(0.000)	(0.010)	(0.169)	(0.003)	(0.000)			
4	224	0.5719	−0.0704	−0.1184	−0.1700	0.5717	−0.0732	−0.1234	−0.1763	0.28	3.69	8.37
		(0.085)	(0.386)	(0.047)	(0.003)	(0.087)	(0.379)	(0.041)	(0.002)			
5	128	0.5657	−0.0635	−0.1337	−0.1741	0.5677	−0.0655	−0.1337	−0.1761	0.32	4.66	10.72
		(0.081)	(0.419)	(0.015)	(0.001)	(0.087)	(0.407)	(0.016)	(0.001)			
6	73	0.5754	−0.0357	−0.1426	−0.1555	0.5807	−0.0333	−0.1356	−0.1484	0.41	5.52	12.10
		(0.068)	(0.689)	(0.012)	(0.011)	(0.067)	(0.705)	(0.017)	(0.014)			
7	41	0.5715	−0.0431	−0.1306	−0.1538	0.5731	−0.0459	−0.1354	−0.1582	0.53	7.46	17.27
		(0.071)	(0.589)	(0.011)	(0.002)	(0.073)	(0.568)	(0.008)	(0.002)			
8	21	0.5603	−0.0483	−0.1235	−0.1217	0.5643	−0.0524	−0.1253	−0.1245	0.75	9.45	19.04
		(0.074)	(0.565)	(0.034)	(0.055)	(0.077)	(0.535)	(0.034)	(0.051)			
9	9	0.4763	−0.1308	−0.2056	−0.2040	0.4923	−0.1230	−0.1982	−0.1980	1.09	14.63	29.08
		(0.127)	(0.116)	(0.000)	(0.002)	(0.121)	(0.145)	(0.001)	(0.003)			
10 (Small)	2	0.3794	−0.2272	−0.2858	−0.2820	0.4158	−0.2005	−0.2733	−0.2651	2.04	21.57	41.86
		(0.224)	(0.004)	(0.000)	(0.000)	(0.192)	(0.016)	(0.000)	(0.000)			

with decreasing fund size. The relation is strictly monotonic and accelerating such that the smallest 10% of funds show by far the highest disappearance rates.

To gain further insight into the determinants of fund disappearance, we use probit models analog to Brown and Goetzmann (1995) to analyze the odds of fund disappearance as a function of yearly (lagged) fund returns, yearly (lagged) fund flows, fund size, expense ratios, and age. Panel I of Table VIII shows results on the models originally used by Brown and Goetzmann (1995). Our results largely confirm theirs by showing that higher returns, size, age, and flows significantly decrease the odds of fund disappearance while higher expense ratios significantly increase the odds of fund disappearance (e.g., Dukes et al., 2006; Gil-Bazo and Ruiz-Verdú, 2009). Also similar to Brown and Goetzmann (1995), we find no economically relevant impact on the interaction between fund size and return on fund disappearance.

In Panel II of Table VIII, we use the same basic model specifications but a different approach to capture the combined impact of returns and size on fund disappearance. To model this interdependency we include dummy variables indicating whether a fund is large (upper third) or small (lower third) in a given month. For Model 5 we obtain an additional negative impact of the relative return ($t-1y$) on the disappearance of small funds which is significantly different from zero. In contrast, we do not find such an additional impact for large funds. Thus, we conclude that the performance-survival relation is much stronger for smaller funds. Model 6 shows that similar relations hold for both relative returns ($t-1y$) and ($t-2y$). Again, for small funds the coefficients are negative and significantly different from zero while the coefficients for large funds are insignificant.

The story behind the performance-survival relation being stronger for smaller funds could be grounded on fund companies maximizing revenues. As the management fee is the primary source of a fund company's revenues, it seems rational to maintain a large fund even if it under-performs in order to retain the fund's assets. At the same time it certainly is rational to close a small under-performing fund which makes only small contribution to the company's revenues in order to dispose of the poor track record. Supporting evidence comes from Gil-Bazo and Ruiz-Verdú (2009) who find that there are significant differences in performance sensitivity of different fund investors. Following their argument, performance sensitive (or sophisticated) investors withdraw their money from under-performing funds while performance insensitive (or unsophisticated) investors – who react to marketing and public awareness rather than to performance – do not. If unsophisticated investors buy larger funds with larger marketing budgets and sophisticated investors buy smaller and rather unknown funds, withdrawals following poor performance should be greater with smaller funds. Moreover, Gil-Bazo and Ruiz-Verdú (2009) find that unsophisticated investors can even be charged higher management fees without risking significant outflows.

Table VIII. Probit analysis of fund disappearance

This table shows results of a probit analysis of fund disappearance between 01/1996 and 12/2006 based on a pooled set of non-overlapping yearly observations. For non-survivors the month of reference for fund observations is the individual date of disappearance (e.g., 05/2000). In all years prior to disappearance these funds count as not disappeared, but the individual month of reference is kept (e.g., 05/1999, 05/1998, etc.). For survivors the month of reference in each year is December. In total, we obtain 15,664 observations, of which 1,198 (7.6%) are fund disappearances. Relative returns represent the total fund return less the average return of all funds. Relative new money is the percentage increase in total net assets of a fund less the average percentage increase in total net assets of all funds. Relative size is the size of a fund less the average size of all funds. Age represents the number of months since fund inception. $(t-1y)$ indicates the 1-year-period prior to the month of reference, $(t-2y)$ the 1-year-period prior to $(t-1y)$, and $(t-1m)$ the month before the month of reference, etc. Panel I shows results for the Brown and Goetzmann (1995) model specifications. Panel II reports results for alternative model specifications where Large (Small) is a dummy variable indicating that a fund belongs to the upper (lower) third with respect to size in $(t-1m)$. p-values are computed using two-sided t-tests for regression coefficients and are based on HAC-consistent covariances (Newey and West, 1987). The McFadden (Nagelkerke) R^2 measures the model fit using the difference in maximum log-likelihood (likelihood) between a specified model and a null model.

	Model 1	p-value	Model 2	p-value	Model 3	p-value	Model 4	p-value
Panel I. Brown and Goetzmann (1995) model specifications								
Intercept	−1.6843	0.000	−1.8039	0.000	−1.7825	0.000	−1.7870	0.000
Relative return $(t-1y)$	−1.2357	0.000	−1.7321	0.000			−1.6947	0.000
Relative return $(t-2y)$			−1.3632	0.000			−1.3420	0.000
Relative return $(t-3y)$							−0.7802	0.000
Relative new money $(t-1y)$			0.0004	0.035	0.0006	0.000		
Relative new money $(t-2y)$			−0.0001	0.193	−0.0002	0.012		
Relative new money $(t-3y)$					−0.0001	0.046		
Relative size $(t-1m)$	−0.0005	0.000	−0.0006	0.000	−0.0007	0.000	−0.0006	0.000
Expense ratio $(t-1m)$	2.9810	0.023	2.4130	0.030	3.4696	0.019	2.1007	0.058
Age $(t-1m)$	−0.0007	0.000	−0.0004	0.048	−0.0007	0.011	−0.0005	0.007
Relative return : Age			−0.0003	0.759			0.0000	0.992
Relative return : Relative size			−0.0007	0.084			−0.0005	0.286
Age : Relative size			0.0000	0.000	0.0000	0.050	0.0000	0.000
Relative NM : Age					0.0000	0.001		
Relative NM : Relative size					0.0000	0.000		
McFadden R^2 (adj.)	0.070		0.094		0.063		0.100	
Nagelkerke R^2	0.090		0.122		0.083		0.129	

(*continued*)

Table VIII. (Continued)

	Model 5	p-value	Model 6	p-value	Model 7	p-value	Model 8	p-value
Panel II. Alternative model specifications								
Intercept	−1.6799	0.000	−1.6850	0.000	−1.7422	0.000	−1.6680	0.000
Relative return ($t-1y$)	−0.9099	0.000	−0.9271	0.000			−1.0326	0.000
Relative return ($t-2y$)			−0.9291	0.000			−0.9607	0.000
Relative return ($t-3y$)							−0.3586	0.027
Relative new money ($t-1y$)			0.0003	0.197	0.0007	0.000		
Relative new money ($t-2y$)			−0.0002	0.058	−0.0002	0.006		
Relative new money ($t-3y$)					−0.0001	0.074		
Relative size ($t-1m$)	−0.0005	0.000	−0.0004	0.000	−0.0006	0.000	−0.0004	0.000
Expense ratio ($t-1m$)	2.7860	0.021	2.0823	0.033	3.4060	0.020	1.5019	0.113
Age ($t-1m$)	−0.0007	0.000	−0.0007	0.000	−0.0007	0.000	−0.0007	0.000
Relative return ($t-1y$) : Large	0.3040	0.202	0.1641	0.548			0.2716	0.315
Relative return ($t-1y$) : Small	−1.0113	0.001	−0.7366	0.016			−0.7778	0.014
Relative return ($t-2y$) : Large			−0.1712	0.520			−0.0250	0.923
Relative return ($t-2y$) : Small			−0.9719	0.001			−0.8654	0.005
Relative return ($t-3y$) : Large							−0.6668	0.015
Relative return ($t-3y$) : Small							−0.7697	0.010
Relative NM ($t-1y$) : Large					−0.0011	0.000		
Relative NM ($t-1y$) : Small					0.0000	0.989		
McFadden R^2 (adj.)	0.072		0.094		0.063		0.102	
Nagelkerke R^2	0.094		0.122		0.082		0.132	

Model 7 shows results on the flow-survival relation. These could be explained by the phenomenon of inflows following superior performance (e.g., Berk and Green, 2004; Edelen et al., 2007; Pollet and Wilson, 2008). The reaction to these inflows is a severe reduction in performance due to decreasing returns to scale and problems in efficiently allocating the new money (e.g., Barras et al., 2010). Therefore, relative inflows (Relative new money ($t - 1y$)) increase the odds of fund disappearance because of poor after-flow performance.[18] For large funds, however, relative inflows (Relative NM ($t - 1y$): Large) decrease the odds of fund disappearance. A rationale for this is that for large funds positive relative new money means high absolute new money, such that large funds are maintained due to the same reason as described above – to generate revenues through management fees.

In Model 8, which includes the relative return ($t - 3y$), we obtain negative coefficients for large and small funds. This means that relative returns lagged three years exhibit an additional impact on the disappearance of large and small funds compared to the middle third of the funds. In general, the results in Panel II of Table VIII explicitly show that the relation between performance and survival is stronger in small funds than it is in large funds.

4.3 PERFORMANCE OF NON-SURVIVORS AND NEW FUNDS

The performance of non-survivors is the main driver of survivorship bias. Therefore, we analyze performance patterns of non-survivors in more detail to get a deeper economic understanding of survivorship bias. Moreover, the performance of new funds seems to be an important issue because fund incubation could explain some of the differences we found between survivorship bias estimates, especially between equal-weighted end-of-sample and full-data survivors.

To show the properties of non-survivors, Panel I of Table IX shows average TNA of different non-survivor fund groups, both in the full period and in different time frames before fund disappearance. The results clearly show non-survivors to be very small and shrinking through the last four years before disappearing. These results are most pronounced for liquidated funds as these are distinctly smaller than non-liquidated funds. Moreover, Panels II–IV of Table IX report performance differences between non-survivors (all, liquidated only, and non-liquidated only) and end-of-sample survivors, both in the full period (e.g., Deaves, 2004) and in different time frames before fund disappearance (e.g., Blake and Timmermann, 1998). In Panel II, the full period performance difference between all non-survivors and end-of-sample survivors is significantly different from zero and economically

[18] In Model 7, lagged returns are not included as explanatory variables. This means that the influence of the flow variables cannot be seen as independent from returns. In contrast, results of Model 6, which includes both flow and return variables, do not show such an economically significant impact of flows on the odds of fund disappearance.

Table IX. Performance differences of non-survivors and end-of-sample survivors, and fund size

This table reports the average fund size (Panel I) of non-survivors (all, liquidated, non-liquidated) and performance differences (Panels II–IV) between non-survivors and end-of-sample survivors in the period from 02/1993 through 11/2002. Results are presented for the "full period" and for different time frames before fund disappearance (e.g., "Last" stands for the last year before disappearance, "2nd" for the second-to-last year before disappearance, etc.). All performance differences are based on the time series of differences between a respective non-survivor portfolio and the end-of-sample survivor portfolio (e.g., equal-weighted non-survivors less equal-weighted end-of-sample). Value-weighted returns are weighted by beginning of month TNA. Performance differences are denoted in percentage points per month. Reported p-values are computed using two-sided t-tests for means and two-sided t-test for regression coefficients, respectively. p-values for regression coefficients are based on HAC-consistent covariances (Newey and West, 1987).

Panel I. Fund size (Mio. US$)

		Year before fund disappearance				
	Full period	Higher than 4th	4th	3rd	2nd	Last
All non-survivors (2,323)	164	367	140	88	58	49
Liquidated (927)	81	150	55	45	27	22
Non-liquidated (1,396)	214	552	213	115	75	62

	Full period		Year before fund disappearance: Higher than 4th		4th		3rd		2nd		Last	
	Measure	p-value	Measure	p-value	Measure	p-value	Measure	p-value	Measure	p-value	Measure	p-value
Panel II. Performance differences: all 2,323 non-survivors (%)												
Equal-weighted												
MER	−0.3610	0.000	−0.2256	0.001	−0.2054	0.001	−0.3660	0.000	−0.5993	0.000	−0.5435	0.000
1F-Alpha	−0.3842	0.001	−0.2619	0.001	−0.2247	0.001	−0.3932	0.001	−0.6269	0.001	−0.5387	0.001
3F-Alpha	−0.3348	0.001	−0.1843	0.001	−0.1685	0.001	−0.3139	0.001	−0.5897	0.001	−0.5514	0.001
4F-Alpha	−0.3067	0.000	−0.1536	0.001	−0.2072	0.001	−0.3401	0.001	−0.5137	0.001	−0.4553	0.001
Value-weighted												
MER	−0.2628	0.000	−0.1919	0.010	−0.2953	0.019	−0.4465	0.001	−0.5414	0.000	−0.4294	0.000
1F-Alpha	−0.2902	0.001	−0.2288	0.001	−0.3175	0.002	−0.4917	0.001	−0.5763	0.001	−0.4319	0.001
3F-Alpha	−0.2248	0.001	−0.1573	0.001	−0.2746	0.002	−0.3754	0.001	−0.5486	0.001	−0.4474	0.001
4F-Alpha	−0.2561	0.001	−0.1750	0.001	−0.3459	0.001	−0.4212	0.001	−0.5138	0.001	−0.3524	0.001

Panel III. Performance differences: 927 liquidated funds (%)												
Equal-weighted												
MER	−0.3824	0.000	−0.2631	0.008	−0.3302	0.000	−0.3567	0.000	−0.5687	0.000	−0.6691	0.000
1F-Alpha	−0.3865	0.000	−0.3041	0.012	−0.3235	0.000	−0.3582	0.000	−0.5720	0.000	−0.6338	0.000
3F-Alpha	−0.3487	0.000	−0.2050	0.005	−0.2740	0.000	−0.2984	0.000	−0.5649	0.000	−0.6793	0.000
4F-Alpha	−0.2969	0.000	−0.1593	0.064	−0.2470	0.001	−0.2922	0.000	−0.4483	0.000	−0.5452	0.000
Value-weighted												
MER	−0.2532	0.000	−0.1777	0.083	−0.2768	0.022	−0.5732	0.001	−0.3590	0.007	−0.6358	0.000
1F-Alpha	−0.2779	0.000	−0.2264	0.033	−0.2695	0.027	−0.6126	0.000	−0.3594	0.010	−0.6373	0.001
3F-Alpha	−0.2260	0.000	−0.1662	0.080	−0.2323	0.081	−0.4751	0.000	−0.4327	0.001	−0.5685	0.001
4F-Alpha	−0.2061	0.000	−0.0748	0.541	−0.3069	0.012	−0.3206	0.015	−0.3469	0.005	−0.6265	0.000
Panel IV. Performance differences: 1,396 non-liquidated funds (%)												
Equal-weighted												
MER	−0.3557	0.000	−0.2176	0.000	−0.1527	0.081	−0.3899	0.000	−0.6196	0.000	−0.4865	0.000
1F-Alpha	−0.3908	0.000	−0.2451	0.000	−0.1883	0.077	−0.4341	0.000	−0.6603	0.000	−0.4970	0.000
3F-Alpha	−0.3341	0.000	−0.1871	0.000	−0.1216	0.198	−0.3389	0.000	−0.6058	0.000	−0.4913	0.000
4F-Alpha	−0.3208	0.000	−0.1754	0.000	−0.2081	0.003	−0.3821	0.000	−0.5490	0.000	−0.4127	0.000
Value-weighted												
MER	−0.2638	0.000	−0.2189	0.003	−0.2829	0.044	−0.4173	0.006	−0.5865	0.000	−0.4168	0.000
1F-Alpha	−0.2917	0.000	−0.2509	0.002	−0.3107	0.072	−0.4628	0.002	−0.6259	0.000	−0.4184	0.000
3F-Alpha	−0.2235	0.000	−0.1758	0.005	−0.2649	0.140	−0.3537	0.007	−0.5745	0.000	−0.4429	0.000
4F-Alpha	−0.2654	0.000	−0.2286	0.000	−0.3279	0.031	−0.4430	0.000	−0.5494	0.000	−0.3237	0.001

relevant across all measures and weighting schemes. The performance differences range from 266 to 471 basis points per year (22.48 to 38.42 basis points per month), depending on measure and weighting scheme.[19] The remaining columns of Panel II show non-survivors to under-perform compared to end-of-sample survivors regardless of the time frame before disappearance. Their performance decreases almost constantly through the last four years, except for the last year where performance surprisingly increases.[20] In their second-to-last year, non-survivors under-perform end-of-sample survivors by 599 to 727 basis points per year (51.37 to 62.69 basis points per month) depending on measure and weighting scheme.

Panels III and IV of Table IX show related performance patterns of liquidated and non-liquidated funds. Generally, there is no clear tendency as to which group performs better, but liquidated funds show by far their worst performance in the last year of existence. In contrast, non-liquidated funds show their worst performance in the second-to-last year and better performance in their last year, which might be related to being merged rather than liquidated. As there are more non-liquidated than liquidated funds and non-liquidated funds are larger, these findings can explain the surprising pattern for all non-survivors described above.

To assess the impact of incubation bias on the results of our empirical analysis, we analyze the properties of new funds. Panel I of Table X shows the average TNA held by new funds in different time frames after fund start. It can be seen that new funds show very low TNA at inception and constantly grow with increasing age. Panel II reports performance differences between new funds and initial funds in the full period and in different time frames after fund start (e.g., Blake and Timmermann, 1998). Regarding performance differences, the first column in Panel II shows that new funds in general slightly under-perform initial funds during the full period but not to a statistically significant degree. When equal-weighted, there is evidence for out-performance of new funds in the first and third year of existence by up to 76 basis points per year (6.29 basis points per month) but the results are not significantly different from zero. In the second year, new funds tend to under-perform initial funds, although again the results are statistically insignificant. With value-weighting, new funds generally under-perform the initial fund portfolio but again without statistical significance.

As described in former sections, incubation seems to have some impact on our results when equal-weighted returns are taken into account. New funds slightly out-perform initial funds in their first and third year, increasing the performance of the equal-weighted end-of-sample portfolio. However, compared to the magnitude

[19] These findings closely correspond to the survivor premium in the UK market reported by Blake and Timmermann (1998), the results reported for US equity funds by Carhart et al. (2002), as well as estimates for the Canadian market reported by Deaves (2004).

[20] The results by Blake and Timmermann (1998) show a similar pattern for "mean abnormal returns" in the six months before fund disappearance.

Table X. Performance differences between new funds and initial funds, and fund size

This table reports the average fund size (Panel I) of new funds and performance differences (Panel II) between new funds and initial funds in the period from 02/1997 through 12/2006. Within this period the table shows size and performance differences for the "full period" as well as measures for different time frames after the fund start (e.g., "1st" stands for the first year of fund existence, etc.). All performance differences are based on the time series of differences between the new funds portfolio and the initial funds portfolio (e.g., equal-weighted new less equal-weighted initial). Value-weighted returns are weighted by the beginning of month TNA. Performance differences are denoted in percentage points per month. Reported p-values are computed using two-sided t-tests for means and two-sided t-test for regression coefficients, respectively. p-values for regression coefficients are based on HAC-consistent covariances (Newey and West, 1987).

Panel I. Fund size (Mio. US$)

		Year of existence				
	Full period	1st	2nd	3rd	4th	5th and higher
New funds	257	51	91	143	195	417

Panel II. Performance differences (%)

	Full period		Year of existence									
			1st		2nd		3rd		4th		5th and higher	
	Measure	p-value	Measure	p-value	Measure	p-value	Measure	p-value	Measure	p-value	Measure	p-value
Equal-weighted												
MER	0.0016	0.967	0.0129	0.211	−0.0531	0.228	0.0629	0.123	0.0235	0.637	−0.0008	0.986
1F-Alpha	−0.0150	0.648	−0.0078	0.911	−0.0728	0.090	0.0614	0.241	0.0121	0.826	−0.0256	0.461
3F-Alpha	−0.0052	0.833	0.0414	0.363	−0.0621	0.085	0.0404	0.416	0.0302	0.545	−0.0217	0.456
4F-Alpha	−0.0213	0.305	0.0239	0.585	−0.0377	0.262	0.0180	0.683	−0.0116	0.773	−0.0474	0.067
Value-weighted												
MER	−0.0051	0.951	−0.0312	0.766	0.0146	0.880	0.0291	0.745	0.0187	0.855	−0.0141	0.857
1F-Alpha	−0.0415	0.539	−0.0638	0.499	−0.0282	0.747	−0.0016	0.984	−0.0192	0.849	−0.0466	0.416
3F-Alpha	−0.0472	0.398	−0.0310	0.656	−0.0102	0.907	−0.0242	0.705	0.0162	0.861	−0.0674	0.137
4F-Alpha	−0.1025	0.016	−0.1027	0.108	−0.0663	0.285	−0.0990	0.105	−0.0421	0.549	−0.1165	0.003

of the under-performance of non-survivors this seems to be of minor importance. Using value-weighted returns, we find no significant or economically relevant out-performance of new funds. On the contrary, new funds tend to under-perform initial funds. With the four-factor alpha, new funds even significantly under-perform in the full period.

5. Summary and Conclusion

Comparing previous studies on mutual funds, it is clear that there is no consistent methodology to estimate survivorship bias. This makes it difficult to interpret and compare the results of different studies correctly. To the best of our knowledge, we are the first to systematically calculate, test, and explain different definitions of the survivorship bias in fund performance. The main methodical differences we identify refer to definitions of survivor and weighting schemes for aggregating fund returns. Analyzing the survivorship bias for US domestic equity mutual fund data, we illuminate this problem by applying different method combinations on a uniform dataset. This allows us to compare the results of different methodologies and show their impact on the magnitude of survivorship bias estimates.

In general, we find positive and statistically significant survivorship bias when ignoring non-survivors, regardless of the methodology applied. Moreover, we show the differences between the methodologies to be systematic. Results for survivorship bias range from 22 to 157 basis points per year. With respect to the weighting scheme applied, equal-weighting yields survivorship bias estimates that are more than twice as high (for full-data conditioning) or five times as high (for end-of-sample conditioning) compared to value-weighting. These differences are both significantly different from zero and economically relevant. This results from non-survivors on average being smaller than survivors. Hence, their influence on the unbiased sample is higher when equal-weighted and smaller when value-weighted.

Concerning the different survivor definitions, the end-of-sample definition shows higher survivorship bias estimates when equal-weighted than the full-data definition. This pattern can partly be attributed to the practice of new fund incubation, which overstates the performance of the equal-weighted end-of-sample portfolio but does not affect the full-data portfolio. With value-weighting the difference almost disappears as the weight of the new funds (both incubated and regularly opened) becomes almost zero. More important, the relation between survival and performance is much stronger in smaller funds. On average, large funds are kept alive when under-performing (presumably) to retain the revenues from management fees while small funds with unsatisfactory performance are more likely to disappear. This explains why end-of-sample funds on average out-perform full-data funds on an equal-weighted basis. With value-weighting, both out-performance of new (incubated) funds and the performance-survival relation are less important due to the small size of new funds and the dominant size of full-data funds within the end-of-sample portfolio.

Our decile-portfolio analysis on the relation between fund size, survival, and performance shows that performance is positively related to size such that large funds out-perform smaller funds. Also, fund disappearance rates increase monotonically with decreasing fund size. As mentioned before, the probit analysis of fund disappearance shows that the relationship between survival and performance is stronger in small funds, which means that the larger a fund, the lower the impact of the performance on its survival.

Analyzing the driver of survivorship bias, we show that non-survivors are smaller and under-perform end-of-sample survivors by 219 to 314 basis points per year. Additionally, we discover size and performance patterns of non-survivors over time. We show fund size to constantly decrease towards disappearance. In general, performance also decreases towards disappearance but slightly increases in the last year of existence. In the second-to-last year, an average non-survivor under-performs survivors by up to 714 basis points per year. Separately analyzing the performance of liquidated and non-liquidated funds as subgroups of non-survivors, we find that only the performance of non-liquidated funds increases in the last year while liquidated funds show their worst performance in their last year of existence.

From these results we can derive several conclusions that might help researchers dealing with survivorship bias. It is important to use value-weighting in order to reduce the bias. This is because non-survivors are generally smaller which means that with equal-weighting their influence is overemphasized. In our case, value-weighting can reduce end-of-sample survivorship bias to one fifth, and with full-data conditioning the bias can be reduced by half. Also, as we find a positive relation between size and performance, equal-weighting does not correctly reflect the performance of the mutual fund market. Further, it is important to learn more about the economic relations reflected by survivorship bias such as the one between fund size, survival, and performance and the composition of different biased samples in order to correctly interpret the results. In our case, these relations explain why the full-data survivor group, which includes mainly large funds, under-performs the end-of-sample survivor group, which includes a very large number of small new funds. Moreover, as there are differences between the performance measures that depend on the factor loadings in combination with the market environment during the sample period, choosing the appropriate model can further reduce survivorship bias.

References

Amin, G. S. and Kat, H. M. (2003) Welcome to the dark side: Hedge fund attrition and survivorship bias over the period 1994–2001, *Journal of Alternative Investment* **Summer**, 53–73.

Arteaga, K. G., Ciccotello, C. S. and Grant, C. T. (1998) New equity funds: Marketing and performance, *Financial Analysts Journal* **54**(6), 43–49.

Barras, L., Scaillet, O. and Wermers, R. (2010) False discoveries in mutual fund performance: Measuring luck in estimated alphas, *Journal of Finance* **65**, 179–216.

Berk, J. B. and Green, R. C. (2004) Mutual fund flows and performance in rational markets, *Journal of Political Economy* **112**, 1269–1295.

Bessler, W., Drobetz, W. and Zimmermann, H. (2009) Conditional performance evaluation for German mutual equity funds, *European Journal of Finance* **15**, 287–316.

Bird, R., Chin, H. and McCrae, M. (1983) The performance of Australian superannuation funds, *Australian Journal of Management* **8**, 49–69.

Blake, C. R., Elton, E. J. and Gruber, M. J. (1993) The performance of bond mutual funds, *Journal of Business* **66**, 371–403.

Blake, D. and Timmermann, A. (1998) Mutual fund performance: Evidence from the UK, *European Finance Review* **2**, 57–77.

Bodson, L., Coën, A. and Hübner, G. (2008) How stable are the major performance measures? *Journal of Performance Measurement* **13**(1), 21–30.

Boynton, W. and Oppenheimer, H. (2006) Anomalies in stock market pricing: Problems in return measurements, *Journal of Business* **79**, 2617–2631.

Brown, S. J. and Goetzmann, W. N. (1995) Performance persistence, *Journal of Finance* **50**, 679–698.

Brown, S. J., Goetzmann, W. N. and Ibbotson, R. G. (1999) Offshore hedge funds: Survival and performance, 1989–95, *Journal of Business* **72**, 91–117.

Brown, S. J., Groetzmann, W. N. and Ross, S. A. (1995) Survival, *Journal of Finance* **50**, 853–873.

Carhart, M. M. (1997) On persistence in mutual fund performance, *Journal of Finance* **52**, 57–82.

Carhart, M. M., Carpenter, J. N., Lynch, A. W. and Musto, D. K. (2002) Mutual fund survivorship, *Review of Financial Studies* **15**, 1439–1463.

Chen, J., Hong, H., Huang, M. and Kubik, J. D. (2004) Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review* **94**, 1276–1302.

Ciccotello, C. and Grant, T. (1996) Equity fund size and growth: Implications for performance and selection, *Financial Services Review* **5**, 1–12.

Coles, J. L., Daniel, N. D. and Nardari, F. (2006) Does the choice of model or benchmark affect inference in measuring mutual fund performance? unpublished working paper, Arizona State University, Purdue University.

Cremers, K. J. M. and Petajisto, A. (2009) How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies* **22**, 3329–3365.

Dahlquist, M., Engström, S. and Söderlind, P. (2000) Performance and characteristics of Swedish mutual funds, *Journal of Financial and Quantitative Analysis* **35**, 409–423.

Deaves, R. (2004) Data-conditioning biases, performance, persistence and flows: The case of Canadian equity funds, *Journal of Banking and Finance* **28**, 673–694.

Droms, W. and Walker, D. (1996) Mutual fund investment performance, *Quarterly Review of Economics and Finance* **36**, 347–363.

Dukes, W. P., English, II, P. C. and Davis, S. M (2006) Mutual fund mortality, 12b-1 fees, and the net expense ratio, *Journal of Financial Research* **29**, 235–352.

Edelen, R. M., Evans, R. and Kadlec, G. B. (2007) Scale effects in mutual fund performance: The role of trading costs, unpublished working paper, Echo Investment Advisors (LLC), University of Virginia, Virginia Tech.

Eling, M. (2008) Does the measure matter in the mutual fund industry? *Financial Analysts Journal* **64**(3), 54–66.

Eling, M. (2009) Does hedge fund performance persist? Overview and new empirical evidence, *European Financial Management* **15**, 362–401.

Elton, E. J., Gruber, M. J. and Blake, C. R. (1996) Survivorship bias and mutual fund performance, *Review of Financial Studies* **9**, 1097–1120.

Evans, R. B. (2006) Does alpha really matter? Evidence from mutual fund incubation, termination and manager change, unpublished working paper, University of Virginia.

Evans, R. B. (2010) Mutual fund incubation, *Journal of Finance* **65**, 1581–1611.

Fama, E. F. and French, K. R. (1993) Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* **33**, 3–56.

Gil-Bazo, J. and Ruiz-Verdú, P. (2009) The relation between price and performance in the mutual fund industry, *Journal of Finance* **64**, 2153–2183.

Gottesmann, A. A. and Morey, M. R. (2007) Predicting emerging market mutual fund performance, *Journal of Investing* **16**, 111–122.

Grinblatt, M. and Titman, S. (1989) Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* **62**, 393–416.

Grinblatt, M. and Titman, S. (1994) A study of monthly mutual fund returns and portfolio performance evaluation techniques, *Journal of Financial and Quantitative Analysis* **29**, 419–444.

Holmes, K. A. and Faff, R. W. (2004) Stability, asymmetry and seasonality of fund performance: An analysis of Australian multisector managed funds, *Journal of Business Finance & Accounting* **31**, 539–578.

Hübner, G. (2007) How do performance measures perform? *Journal of Portfolio Management* **33** (Summer), 64–74.

Indro, D. C., Jiang, C. X., Hu, M. V. and Lee, W. Y. (1999) Mutual fund performance: Does fund size matter? *Financial Analysts Journal* **55**(3), 74–87.

Jensen, M. C. (1968) The performance of mutual funds in the period 1945–1964, *Journal of Finance* **23**, 389–416.

Karoui, A. and Meier, I. (2009) Performance and characteristics of mutual fund starts, *European Journal of Finance* **15**, 487–509.

Liang, B. (2000) Hedge funds: The living and the dead, *Journal of Financial & Quantitative Analysis* **45**, 309–326.

Malkiel, B. G. (1995) Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance* **50**, 549–572.

Malkiel, B. G. and Saha, A. (2005) Hedge funds: Risk and return, *Financial Analysts Journal* **61**(6), 80–88.

Matallín-Sáez, J. C. (2006) Seasonality, market timing and performance amongst benchmarks and mutual fund evaluation, *Journal of Business Finance & Accounting* **33**, 1484–1507.

Newey, W. K. and West, K. D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**, 703–708.

Otten, R. and Bams, D. (2002) European mutual fund performance, *European Financial Management* **8**, 75–101.

Otten, R. and Bams, D. (2004) How to measure mutual fund performance: Economic versus statistical relevance, *Journal of Accounting and Finance* **44**, 203–222.

Pástor, L. and Stambaugh, R. F. (2002) Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* **63**, 315–349.

Pollet, J. M. and Wilson, M. (2008) How does size affect mutual fund behavior? *Journal of Finance* **63**, 2941–2969.

Scholz, H. and Schnusenberg, O. (2009) Ranking of equity mutual funds: The bias in using survivorship bias-free datasets, unpublished working paper, University of Erlangen-Nuernberg and University of North Florida.

ter Horst, J. R., Nijman, T. E. and Verbeek, M. (2001) Eliminating look-ahead bias in evaluating persistence in mutual fund performance, *Journal of Empirical Finance* **8**, 345–373.
ter Horst, J. R. and Verbeek, M. (2007) Fund liquidation, self-selection, and look-ahead bias in the hedge fund industry, *Review of Finance* **11**, 605–632.
Wermers, R. (1997) Momentum investment strategies of mutual funds, performance persistence, and survivorship Bias, unpublished working paper, University of Colorado at Boulder, Boulder.
Zhao, X. (2005) Exit decisions in the U.S. mutual fund industry, *Journal of Business* **78**, 1365–1401.